

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 1999, 15:03:38 ; Search time 1096.08 Seconds  
(without alignments)  
35.993 Million cell updates/sec

Title: US-09-026-400-6

Perfect score: 20

Sequence: 1 gcdatrtgncraanaccc 20

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:\*\*

1: em\_est1:\*\*

2: em\_est2:\*\*

3: em\_est3:\*\*

4: em\_est4:\*\*

5: em\_est5:\*\*

6: em\_est6:\*\*

7: em\_est7:\*\*

8: em\_est8:\*\*

9: em\_est9:\*\*

10: em\_est10:\*\*

11: em\_est11:\*\*

12: em\_est12:\*\*

13: em\_est13:\*\*

14: em\_est14:\*\*

15: em\_est15:\*\*

16: em\_est16:\*\*

17: em\_est17:\*\*

18: em\_est18:\*\*

19: em\_est19:\*\*

20: gb\_est1:\*\*

21: gb\_est2:\*\*

22: gb\_est3:\*\*

23: gb\_est4:\*\*

24: gb\_est5:\*\*

25: gb\_est6:\*\*

26: gb\_est7:\*\*

27: gb\_est8:\*\*

28: gb\_est9:\*\*

29: gb\_est10:\*\*

30: gb\_est11:\*\*

31: gb\_est12:\*\*

32: gb\_est13:\*\*

33: gb\_est14:\*\*

34: gb\_est15:\*\*

35: gb\_est16:\*\*

36: gb\_est17:\*\*

37: gb\_est18:\*\*

38: gb\_est19:\*\*

39: gb\_est20:\*\*

40: gb\_est21:\*\*

41: gb\_est22:\*\*

42: gb\_est23:\*\*

43: gb\_est24:\*\*

44: gb\_est25:\*\*

45: gb\_est26:\*\*

46: gb\_est27:\*\*

47: gb\_est28:\*\*

48: gb\_est29:\*\*

49: gb\_est30:\*\*

50: gb\_est31:\*\*

51: gb\_est32:\*\*

52: em\_est20:\*\*

53: em\_est21:\*\*

54: em\_est22:\*\*  
55: em\_est23:\*\*  
56: em\_est24:\*\*  
57: em\_est25:\*\*  
58: em\_est26:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	14.4	72.0	599	40	AA941765 LD25946.5
2	14.4	72.0	785	40	AA952006 LD29007.5
3	14.4	72.0	649	40	AA990815 LD34717.5
4	14.4	72.0	652	41	AI061712 LD33793.5
5	14.4	72.0	672	44	AI257428 LP05788.5
6	14.4	72.0	649	44	AI259611 LP03065.5
7	14.4	72.0	609	47	AI512710 LD44673.5
8	14.4	72.0	613	47	AI512875 LD44958.5
9	14.4	72.0	658	47	AI515803 LD41473.5
10	14.4	72.0	586	47	AI531887 SD02921.5
11	14.4	72.0	444	48	AI569754 tn05d11.x
12	13.8	69.0	643	33	AA390566 LD09416.5
13	13.8	69.0	721	33	AA439525 LD14116.5
14	13.8	69.0	645	33	AA440182 LD14716.5
15	13.8	69.0	504	33	AA440803 LD15683.5
16	13.8	69.0	414	37	AA680465 SWAMCA304
17	13.8	69.0	587	37	AA695021 GM01953.5
18	13.8	69.0	720	37	AA704069 ag47b03.s
19	13.8	69.0	536	39	AA870850 vq27c01.r
20	13.8	69.0	326	39	AA883627 al57c07.s
21	13.8	69.0	381	40	AA936014 oh34g01.s
22	13.8	69.0	657	40	AA941506 LD25672.5
23	13.8	69.0	577	40	AA951871 LD33724.5
24	13.8	69.0	332	40	AA969205 op51d05.s
25	13.8	69.0	599	41	AI012608 EST207059
26	13.8	69.0	411	43	AI176025 EST219601
27	13.8	69.0	560	45	AI353745 zeh0929.s
28	13.8	69.0	439	46	AA926282 UI-R-AI-e
29	13.8	69.0	439	46	AA955061 UI-R-AI-e
30	13.8	69.0	615	47	AI514859 LD41284.5
31	13.8	69.0	574	47	AI514937 LD46574.5
32	13.8	69.0	581	47	AI516637 LD47712.5
33	13.8	69.0	660	48	AI546381 LD47943.5
34	13.8	69.0	654	48	AI54779 tn30g01.x
35	13.8	69.0	437	48	AI571347 tn45a12.x
36	13.8	69.0	546	48	AI59587 EST251290
37	13.8	69.0	538	48	AI684230 tx80c01.x
38	13.8	69.0	504	50	AI684230 tx80c01.x
39	13.4	67.0	417	21	R11658 yf40c11.s1
40	13.4	67.0	348	22	R80751 y197409.s1
41	13.4	67.0	390	23	R95042 yq44d02.r1
42	13.4	67.0	448	23	R95183 yq45h02.r1
43	13.4	67.0	383	32	AA378645 EST91539
44	13.4	67.0	438	33	AA411566 zv22g01.r
45	13.4	67.0	404	33	AA411567 zv22g01.s

ALIGNMENTS

RESULT 1

AA941765

LOCUS

DEFINITION

ACCESION

NID

VERSION

AA941765

LD25946.5prime LD Drosophila melanogaster embryo p0T2 Drosophila

melanogaster cDNA clone LD25946 5prime, mRNA sequence.

AA941765

93101598

599 bp

mRNA

EST

25-NOV-1998

Fri Oct 1 16:47:14 1999

KEYWORDS  
SOURCE  
ORGANISM  
Drosophila melanogaster  
fruit fly.  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
1 (bases 1 to 599)  
Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,  
Brokstein,P., Lewis,S. and Rubin,G.M.  
BDGP/HMI Drosophila EST Project  
Unpublished (1997)  
On Dec 18, 1997 this sequence version replaced gi:2339750.

Contact: Harvey, D.  
G. M. Rubin-Molecular and Cell Biology  
University of California Berkeley  
539 LSA, Berkeley, CA 94720-3200, USA  
Fax: 510 643 9947  
Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)  
Plate: 259 row: D column: 10  
High quality sequence stop: 553.

FEATURES  
Location/Qualifiers  
1..599

Source  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="LD25946"  
/clone\_lib="LD Drosophila melanogaster embryo pOT2"  
/sex="male and female"  
/dev\_stage="0 to 24 hours mixed stage embryonic"  
/lab\_host="XL1 Blue"  
/note="Organ: embryo; Vector: pOT2; Site\_1: EcoRI; Site\_2:  
XhoI; Sized fractionated cDNAs were directly ligated into  
pOT2."

BASE COUNT 161 a 161 c 142 g 135 t

Query Match 72.0%; Score 14.4; DB 40; Length 599;  
Best Local Similarity 68.4%; Pred. No. 1.5e+02;  
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gcdatrtgncraanacnc 19  
||:|||||:|||||  
Db 543 GCGATGTGCCGAAACGC 561

RESULT 2  
AA952006 785 bp mRNA EST 25-NOV-1998  
LOCUS LD29007.5prime LD Drosophila melanogaster embryo pOT2 Drosophila  
melanogaster cDNA clone LD29007 5prime, mRNA sequence.  
ACCESSION AA952006  
NID 93111819  
VERSION AA952006.1 GI:3111819  
KEYWORDS EST.  
SOURCE fruit fly.

ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
1 (bases 1 to 785)  
Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,  
Brokstein,P., Lewis,S. and Rubin,G.M.  
BDGP/HMI Drosophila EST Project  
Unpublished (1997)  
On Jan 19, 1998 this sequence version replaced gi:2287479.

Contact: Harvey, D.  
G. M. Rubin-Molecular and Cell Biology  
University of California Berkeley  
539 LSA, Berkeley, CA 94720-3200, USA  
Fax: 510 643 9947  
Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)

FEATURES  
source

Plate: 290 row: A column: 7  
High quality sequence stop: 754.  
Location/Qualifiers  
1..785  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/map="846F08"  
/clone="LD29007"  
/clone\_lib="LD Drosophila melanogaster embryo pOT2"  
/sex="male and female"  
/dev\_stage="0 to 24 hours mixed stage embryonic"  
/lab\_host="XL1 Blue"  
/note="Organ: embryo; Vector: pOT2; Site\_1: EcoRI; Site\_2:  
XhoI; Sized fractionated cDNAs were directly ligated into  
pOT2."

BASE COUNT 209 a 223 c 176 g 177 t

Query Match 72.0%; Score 14.4; DB 40; Length 785;  
Best Local Similarity 68.4%; Pred. No. 1.6e+02;  
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gcdatrtgncraanacnc 19  
||:|||||:|||||  
Db 554 GCGATGTGCCGAAACGC 572

RESULT 3  
AA990815 649 bp mRNA EST 24-NOV-1998  
LOCUS LD34717.5prime LD Drosophila melanogaster embryo pOT2 Drosophila  
melanogaster cDNA clone LD34717 5prime, mRNA sequence.  
ACCESSION AA990815  
NID 93177348  
VERSION AA990815.1 GI:3177348  
KEYWORDS EST.  
SOURCE fruit fly.

ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
1 (bases 1 to 649)  
Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,  
Brokstein,P., Lewis,S. and Rubin,G.M.  
BDGP/HMI Drosophila EST Project  
Unpublished (1997)  
On Jan 19, 1998 this sequence version replaced gi:2287188.

Contact: Harvey, D.  
G. M. Rubin-Molecular and Cell Biology  
University of California Berkeley  
539 LSA, Berkeley, CA 94720-3200, USA  
Fax: 510 643 9947  
Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)  
Plate: 347 row: B column: 5  
High quality sequence stop: 529.

FEATURES  
source

Location/Qualifiers  
1..649  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="LD34717"  
/clone\_lib="LD Drosophila melanogaster embryo pOT2"  
/sex="male and female"  
/dev\_stage="0 to 24 hours mixed stage embryonic"  
/lab\_host="XL1 Blue"  
/note="Organ: embryo; Vector: pOT2; Site\_1: EcoRI; Site\_2:  
XhoI; Sized fractionated cDNAs were directly ligated into  
pOT2."

BASE COUNT 173 a 175 c 158 g 143 t

```

Query Match      72.0%; Score 14.4; DB 40; Length 649;
Best Local Similarity 68.4%; Pred. No. 1.5e+02;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gcdatrtgncraanaacnc 19
Db 521 GCGATGTGCCCGAAACGC 539

RESULT 4
AI061712      652 bp mRNA EST 24-NOV-1998
LOCUS LD33793.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
DEFINITION melanogaster cDNA clone LD33793 5prime, mRNA sequence.
ACCESSION AI061712
VERSION 93337628
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 652)
AUTHORS Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
Brokstein,P., Lewis,S. and Rubin,G.M.
JOURNAL BDGP/HMI Drosophila EST Project
UNPUBLISHED (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2152369.

Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 337 row: H column: 9
High quality sequence stop: 576.
Location/Qualifiers
1..652
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/map="21q"
/clone="LD33793"
/clone_lib="LD Drosophila melanogaster embryo pOT2"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="X11 Blue"
/note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pOT2."
XhoI; Sized fractionated cDNAs were directly ligated into
pOT2.

BASE COUNT 177 a 178 c 156 g 141 t
ORIGIN

Query Match      72.0%; Score 14.4; DB 41; Length 652;
Best Local Similarity 68.4%; Pred. No. 1.5e+02;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gcdatrtgncraanaacnc 19
Db 533 GCGATGTGCCCGAAACGC 551

RESULT 5
AI257428      672 bp mRNA EST 17-NOV-1998
LOCUS LP05788.5prime LP Drosophila melanogaster larval-early pupal pOT2
DEFINITION Drosophila melanogaster cDNA clone LP05788 5prime, mRNA sequence.
ACCESSION AI257428
VERSION 93864953
KEYWORDS EST.

```

```

fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 672)
AUTHORS Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
Brokstein,P., Lewis,S. and Rubin,G.M.
JOURNAL BDGP/HMI Drosophila EST Project
UNPUBLISHED (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2150302.

Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 57 row: H column: 4
High quality sequence stop: 640.
Location/Qualifiers
1..672
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/map="6p21.3"
/clone="LP05788"
/clone_lib="LP Drosophila melanogaster larval-early pupal
pOT2"
/sex="male and female"
/dev_stage="larvae-pupae"
/lab_host="DH5-alpha"
/note="Organ: whole body; Vector: pOT2; Site_1: EcoRI;
Site_2: XhoI; Sized fractionated cDNAs were directly
ligated into pOT2. Plasmid cDNA library."

BASE COUNT 174 a 184 c 163 g 151 t
ORIGIN

Query Match      72.0%; Score 14.4; DB 44; Length 672;
Best Local Similarity 68.4%; Pred. No. 1.5e+02;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gcdatrtgncraanaacnc 19
Db 515 GCGATGTGCCCGAAACGC 533

RESULT 6
AI259611      649 bp mRNA EST 17-NOV-1998
LOCUS LP03065.5prime LP Drosophila melanogaster larval-early pupal pOT2
DEFINITION Drosophila melanogaster cDNA clone LP03065 5prime, mRNA sequence.
ACCESSION AI259611
VERSION 93867136
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 649)
AUTHORS Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
Brokstein,P., Lewis,S. and Rubin,G.M.
JOURNAL BDGP/HMI Drosophila EST Project
UNPUBLISHED (1997)
COMMENT On Aug 21, 1998 this sequence version replaced.

Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA

```

Fri Oct 1 16:47:14 1999

us-09-026-400-6.rst

Fax: 510 643 9947  
 Email: <http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu>  
 Plate: 30 row: F column: 5  
 High quality sequence stop: 483.

## FEATURES

source  
 1..649  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone\_lib="LP03065"  
 /clone\_lib="LP Drosophila melanogaster larval-early pupal  
 pOT2"  
 /sex="male and female"  
 /dev\_stage="larvae-pupae"  
 /lab\_host="DHS-alpha"  
 /note="Organ: whole body; Vector: pOT2; Site\_1: EcoRI;  
 Site\_2: XhoI; Sized fractionated cDNAs were directly  
 ligated into pOT2. Plasmid cDNA library."

BASE COUNT 177 a 177 c 153 g 142 t  
 ORIGIN

Query Match 72.0%; Score 14.4; DB 44; Length 649;  
 Best Local Similarity 68.4%; Pred. No. 1.5e+02;  
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 gdatrtgncraaacnc 19  
 ||:|||| ||:||||  
 Db 548 GCGATGTGCCGAAAACGC 566

## RESULT 7

AI512710 609 bp mRNA EST 16-MAR-1999  
 LOCUS LD44673.5prime LD Drosophila melanogaster embryo pOT2 Drosophila  
 DEFINITION melanogaster cDNA clone LD44673 5prime, mRNA sequence.  
 ACCESSION AI512710  
 NID 94422128  
 VERSION AI512710.1 GI:4422128  
 SOURCE EST.  
 ORGANISM fruit fly.  
 Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 609)  
 AUTHORS Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,  
 Brokstein,P., Lewis,S. and Rubin,G.M.  
 TITLE BDGP/HMMI Drosophila EST Project  
 JOURNAL Unpublished (1997)  
 COMMENT On May 18, 1998 this sequence version replaced gi:3136738.

Contact: Harvey, D.  
 G. M. Rubin-Molecular and Cell Biology  
 University of California Berkeley  
 539 LSA, Berkeley, CA 94720-3200, USA  
 Fax: 510 643 9947  
 Email: <http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu>  
 Plate: 446 row: G column: 1  
 High quality sequence stop: 513.

## FEATURES

source  
 1..609  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone\_lib="LD44673"  
 /clone\_lib="LD Drosophila melanogaster embryo pOT2"  
 /sex="male and female"  
 /dev\_stage="0 to 24 hours mixed stage embryonic"  
 /lab\_host="Xli Blue"  
 /note="Organ: embryo; Vector: pOT2; Site\_1: EcoRI; Site\_2:  
 XhoI; Sized fractionated cDNAs were directly ligated into  
 pOT2."

BASE COUNT 165 a 166 c 148 g 130 t

## ORIGIN

Query Match 72.0%; Score 14.4; DB 47; Length 609;  
 Best Local Similarity 68.4%; Pred. No. 1.5e+02;  
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 gdatrtgncraaacnc 19  
 ||:|||| ||:||||  
 Db 523 GCGATGTGCCGAAAACGC 541

## RESULT 8

AI512875 613 bp mRNA EST 16-MAR-1999  
 LOCUS LD44958.5prime LD Drosophila melanogaster embryo pOT2 Drosophila  
 DEFINITION melanogaster cDNA clone LD44958 5prime, mRNA sequence.  
 ACCESSION AI512875  
 NID 94422293  
 VERSION AI512875.1 GI:4422293  
 KEYWORDS EST.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 613)  
 AUTHORS Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,  
 Brokstein,P., Lewis,S. and Rubin,G.M.  
 TITLE BDGP/HMMI Drosophila EST Project  
 JOURNAL Unpublished (1997)  
 COMMENT On May 18, 1998 this sequence version replaced gi:3136948.

Contact: Harvey, D.  
 G. M. Rubin-Molecular and Cell Biology  
 University of California Berkeley  
 539 LSA, Berkeley, CA 94720-3200, USA  
 Fax: 510 643 9947  
 Email: <http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu>  
 Plate: 449 row: E column: 10  
 High quality sequence stop: 581.

## FEATURES

source  
 1..613  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone\_lib="LD44958"  
 /clone\_lib="LD Drosophila melanogaster embryo pOT2"  
 /sex="male and female"  
 /dev\_stage="0 to 24 hours mixed stage embryonic"  
 /lab\_host="Xli Blue"  
 /note="Organ: embryo; Vector: pOT2; Site\_1: EcoRI; Site\_2:  
 XhoI; Sized fractionated cDNAs were directly ligated into  
 pOT2."

BASE COUNT 166 a 168 c 143 g 136 t  
 ORIGIN

Query Match 72.0%; Score 14.4; DB 47; Length 613;  
 Best Local Similarity 68.4%; Pred. No. 1.5e+02;  
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 gdatrtgncraaacnc 19  
 ||:|||| ||:||||  
 Db 554 GCGATGTGCCGAAAACGC 572

## RESULT 9

AI515803 658 bp mRNA EST 16-MAR-1999  
 LOCUS LD44475.5prime LD Drosophila melanogaster embryo pOT2 Drosophila  
 DEFINITION melanogaster cDNA clone LD44475 5prime, mRNA sequence.  
 ACCESSION AI515803  
 NID 94418903

```

VERSION      AI515803.1  GI:4418903
KEYWORDS     EST.
SOURCE       Drosophila melanogaster
ORGANISM     fruit fly.
             Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
             Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
             Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1 (bases 1 to 658)
AUTHORS      Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
             Brokstein,P., Lewis,S. and Rubin,G.M.
TITLE        BDGP/HMI Drosophila EST Project
JOURNAL      Unpublished (1997)
COMMENT      On Mar 10, 1998 this sequence version replaced gi:2948693.

Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 414 row: G column: 3
High quality sequence stop: 464.
Location/Qualifiers
  1..658
  /organism="Drosophila melanogaster"
  /db_xref="taxon:7227"
  /clone="LD41475"
  /clone_lib="LD Drosophila melanogaster embryo pOT2"
  /sex="male and female"
  /dev_stage="0 to 24 hours mixed stage embryonic"
  /lab_host="XLI Blue"
  /note="Organ: embryo; Vector: pOT2; Site.1: EcoRI; Site.2:
XhoI; Sized fractionated cDNAs were directly ligated into
pOT2."
BASE COUNT   172 a 178 c 162 g 146 t
ORIGIN

Query Match      72.0%; Score 14.4; DB 47; Length 658;
Best Local Similarity 68.4%; Pred. No. 1.5e-02;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY  1  gdatrtgncraaacnc 19
      ||:|| ||:|| |||
Db  509  GCGATGTCGCCGAAACGC 527

RESULT 10
AI531687
LOCUS      SD02921.5prime SD Drosophila melanogaster Schneider L2 cell culture
DEFINITION  pOT2 Drosophila melanogaster cDNA clone SD02921 5prime, mRNA
sequence.
ACCESSION  AI531687
NID        94445822
VERSION    AI531687.1  GI:4445822
KEYWORDS   EST.
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 586)
AUTHORS    Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
TITLE      BDGP/HMI Drosophila EST Project
JOURNAL    Unpublished (1997)
COMMENT    On Mar 10, 1998 this sequence version replaced gi:2948379.

Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA

```

```

Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 29 row: B column: 9
High quality sequence stop: 477.
Location/Qualifiers
  1..586
  /organism="Drosophila melanogaster"
  /db_xref="taxon:7227"
  /map="11q13.1"
  /clone="SD02921"
  /clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture pOT2"
  /lab_host="DH5-alpha"
  /note="Vector: pOT2; Site.1: EcoRI; Site.2: XhoI; Sized
fractionated cDNAs were directly ligated into pOT2.
Plasmid cDNA library."
BASE COUNT   158 a 160 c 138 g 130 t
ORIGIN

Query Match      72.0%; Score 14.4; DB 47; Length 586;
Best Local Similarity 68.4%; Pred. No. 1.5e-02;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY  1  gcdattgncraaacnc 19
      ||:|| || ||:|| |||
Db  543  GCGATGTCGCCGAAACGC 561

RESULT 11
AI569754
LOCUS      tn05g11.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone
DEFINITION  mRNA sequence.
ACCESSION  AI569754
NID        94533128
VERSION    AI569754.1  GI:4533128
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 444)
AUTHORS    NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGP), Tumor Gene Index
TITLE      Unpublished (1998)
JOURNAL    On Mar 10, 1998 this sequence version replaced gi:2948525.
COMMENT    Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LINL at:
            www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 472 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 430.
Location/Qualifiers
  1..444
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="IMAGE:2166741"
  /clone_lib="NCI_CGAP_Brn25"
  /tissue_type="anaplastic oligodendroglioma"

```

/lab\_host="DH10B"  
 /note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAATGGAGCGCGCATAGGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."  
 BASE COUNT 97 a 129 c 130 g 87 t 1 others  
 ORIGIN

Query Match 72.0%; Score 14.4; DB 48; Length 444;  
 Best Local Similarity 68.4%; Pred. No. 1.4e+02;  
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 cdatttgnccraanacc 20  
 ||:|||||:|||||  
 Db 201 CAATGTGCGCGAAACACC 219

RESULT 12  
 AA390566/c 643 bp mRNA EST 28-NOV-1998  
 LOCUS LD09416.5prime LD Drosophila melanogaster embryo Bluescript  
 DEFINITION Drosophila melanogaster cDNA clone LD09416 5prime, mRNA sequence.

ACCESSION AA390566  
 NID Q2043560  
 VERSION AA390566.1 GI:2043560  
 KEYWORDS EST.  
 SOURCE fruit fly.

ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 643)  
 AUTHORS Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein,P., Lewis,S. and Rubin,G.M.  
 TITLE BDGP/HMI Drosophila EST Project  
 JOURNAL Unpublished (1997)  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1395289.

Contact: Harvey, D.  
 G. M. Rubin-Molecular and Cell Biology  
 University of California Berkeley  
 539 LSA, Berkeley, CA 94720-3200, USA  
 Fax: 510 643 9947  
 Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu  
 Plate: 94 row: B column: 4  
 High quality sequence stop: 578.  
 Location/Qualifiers  
 1. 643  
 /organism="Drosophila melanogaster"  
 /db\_xref="BDGP\_EST:BDcln008715"  
 /db\_xref="taxon:7227"  
 /clone="LD09416"  
 /clone\_lib="LD Drosophila melanogaster embryo Bluescript"  
 /sex="male and female"  
 /dev\_stage="0 to 24 hours mixed stage embryonic"  
 /lab\_host="SOLR"  
 /note="Organ: brain; Vector: Bluescript SK; Site\_1: EcoRI; Site\_2: XhoI; Constructed using Stratagene ZAP-cDNA Synthesis kit. Oligo dT-primed and directionally cloned at EcoRI and XhoI in Bluescript SK(+/-)"

BASE COUNT 168 a 176 c 155 g 144 t  
 ORIGIN

Query Match 69.0%; Score 13.8; DB 33; Length 643;  
 Best Local Similarity 65.0%; Pred. No. 3.4e+02;  
 Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 gcdatttgnccraanacc 20  
 ||:|||||:|||||  
 Db 289 GCAATGTGCGCGAAATCTCC 270

RESULT 13  
 AA439525/c 721 bp mRNA EST 28-NOV-1998  
 LOCUS LD14791.5prime LD Drosophila melanogaster embryo Bluescript  
 DEFINITION Drosophila melanogaster cDNA clone LD14791 5prime, mRNA sequence.

ACCESSION AA439525  
 NID G2151406  
 VERSION AA439525.1 GI:2151406  
 KEYWORDS EST.  
 SOURCE fruit fly.

ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 721)  
 AUTHORS Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein,P., Lewis,S. and Rubin,G.M.  
 TITLE BDGP/HMI Drosophila EST Project  
 JOURNAL Unpublished (1997)  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1394108.

Contact: Harvey, D.  
 G. M. Rubin-Molecular and Cell Biology  
 University of California Berkeley  
 539 LSA, Berkeley, CA 94720-3200, USA  
 Fax: 510 643 9947  
 Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu  
 Plate: 141 row: B column: 4  
 High quality sequence stop: 613.  
 Location/Qualifiers  
 1. 721  
 /organism="Drosophila melanogaster"  
 /db\_xref="BDGP\_EST:BDcln013323"  
 /db\_xref="taxon:7227"  
 /clone="LD14791"  
 /clone\_lib="LD Drosophila melanogaster embryo Bluescript"  
 /sex="male and female"  
 /dev\_stage="0 to 24 hours mixed stage embryonic"  
 /lab\_host="SOLR"  
 /note="Organ: embryo; Vector: Bluescript SK; Site\_1: EcoRI; Site\_2: XhoI; Constructed using Stratagene ZAP-cDNA Synthesis kit. Oligo dT-primed and directionally cloned at EcoRI and XhoI in Bluescript SK(+/-)"

BASE COUNT 187 a 198 c 179 g 157 t  
 ORIGIN

FEATURES  
 source

Query Match 69.0%; Score 13.8; DB 33; Length 721;  
 Best Local Similarity 65.0%; Pred. No. 3.4e+02;  
 Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 gcdatttgnccraanacc 20  
 ||:|||||:|||||  
 Db 260 GCAATGTGCGCGAAATCTCC 241

RESULT 14  
 AA440182/c 645 bp mRNA EST 28-NOV-1998  
 LOCUS AA440182.1 GI:2151920  
 DEFINITION Drosophila melanogaster cDNA clone LD14791 5prime, mRNA sequence.

ACCESSION AA440182  
 NID G2151920  
 VERSION AA440182.1 GI:2151920  
 KEYWORDS EST.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster

REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 AUTHORS 1 (bases 1 to 645)  
 TITLE Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein,P., Lewis,S. and Rubin,G.M.  
 JOURNAL BDGP/HMI Drosophila EST Project  
 COMMENT Unpublished (1997)  
 On Sep 12, 1996 this sequence version replaced gi:1394984.

Contact: Harvey, D.  
 G. M. Rubin-Molecular and Cell Biology  
 University of California Berkeley  
 539 LSA, Berkeley, CA 94720-3200, USA  
 Fax: 510 643 9947  
 Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)  
 Plate: 147 row: H column: 7  
 High quality sequence stop: 472.

FEATURES  
 Source Location/Qualifiers  
 1..845  
 /organism="Drosophila melanogaster"  
 /db\_xref="BDGP\_EST:BDcIn013974"  
 /db\_xref="taxon:7227"  
 /clone="LD14791"  
 /clone\_lib="LD Drosophila melanogaster embryo Bluescript"  
 /sex="male and female"  
 /dev\_stage="0 to 24 hours mixed stage embryonic"  
 /lab\_host="SOLR"  
 /note="Organ: embryo; Vector: Bluescript SK; Site:1;  
 ECORI; Site:2: XhoI; Constructed using Stratagene ZAP-cDNA  
 Synthesis kit. Oligo dt-primed and directionally cloned at  
 ECORI and XhoI in Bluescript SK(+/-)"  
 BASE COUNT 168 a 170 c 158 g 149 t  
 ORIGIN

Query Match 59.0%; Score 13.8; DB 33; Length 645;  
 Best Local Similarity 65.0%; Pred. No. 3.4e+02;  
 Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 gcdatttgcraanacncc 20  
 ||:|||| ||:| | |  
 Db 303 GCAATGTGCCGAATCTCC 284

RESULT 15  
 AA440803/C  
 LOCUS AA440803 504 bp mRNA EST 28-NOV-1998  
 DEFINITION LD15683.Sprime LD Drosophila melanogaster embryo Bluescript  
 Drosophila melanogaster cDNA clone LD15683 Sprime, mRNA sequence.  
 ACCESSION AA440803  
 NID 92152681  
 VERSION AA440803.1 GI:2152681  
 KEYWORDS EST.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 504)  
 AUTHORS Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,  
 Brokstein,P., Lewis,S. and Rubin,G.M.  
 TITLE BDGP/HMI Drosophila EST Project  
 JOURNAL Unpublished (1997)  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1394584.

Contact: Harvey, D.  
 G. M. Rubin-Molecular and Cell Biology  
 University of California Berkeley  
 539 LSA, Berkeley, CA 94720-3200, USA  
 Fax: 510 643 9947  
 Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)  
 Plate: 156 row: G column: 11

FEATURES  
 Source Location/Qualifiers  
 1..504  
 /organism="Drosophila melanogaster"  
 /db\_xref="BDGP\_EST:BDcIn014830"  
 /db\_xref="taxon:7227"  
 /clone="LD15683"  
 /clone\_lib="LD Drosophila melanogaster embryo Bluescript"  
 /sex="male and female"  
 /dev\_stage="0 to 24 hours mixed stage embryonic"  
 /lab\_host="SOLR"  
 /note="Organ: embryo; Vector: Bluescript SK; Site:1;  
 ECORI; Site:2: XhoI; Constructed using Stratagene ZAP-cDNA  
 Synthesis kit. Oligo dt-primed and directionally cloned at  
 ECORI and XhoI in Bluescript SK(+/-)"  
 BASE COUNT 129 a 128 c 124 g 123 t  
 ORIGIN

Query Match 59.0%; Score 13.8; DB 33; Length 504;  
 Best Local Similarity 65.0%; Pred. No. 3.2e+02;  
 Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 gcdatttgcraanacncc 20  
 ||:|||| ||:| | |  
 Db 296 GCAATGTGCCGAATCTCC 277

Search completed: October 1, 1999, 15:03:42  
 Job time: 4386 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: October 1, 1999, 15:34:44 ; Search time 1619.02 Seconds  
(without alignments)  
39.287 Million cell updates/sec

Title: US-09-026-400-6  
Perfect score: 20  
Sequence: 1 gcdatrtgncraanacnc 20

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl.\*

1: gb\_ba1.\*

2: gb\_ba2.\*

3: gb\_om.\*

4: gb\_ov.\*

5: gb\_pat.\*

6: gb\_ph.\*

7: gb\_pl1.\*

8: gb\_pl2.\*

9: gb\_pr1.\*

10: gb\_pr2.\*

11: gb\_pr3.\*

12: gb\_ro.\*

13: gb\_st.\*

14: gb\_sts.\*

15: gb\_sy.\*

16: gb\_un.\*

17: gb\_vi.\*

18: em\_fun.\*

19: em\_htg.\*

20: em\_hum1.\*

21: em\_hum2.\*

22: em\_in.\*

23: em\_om.\*

24: em\_or.\*

25: em\_ov.\*

26: em\_pat.\*

27: em\_ph.\*

28: em\_pl.\*

29: em\_ro.\*

30: em\_sts.\*

31: em\_sy.\*

32: em\_un.\*

33: em\_vi.\*

34: gb\_htg1.\*

35: gb\_htg2.\*

36: gb\_in1.\*

37: gb\_in2.\*

38: em\_ba1.\*

39: em\_ba2.\*

40: em\_hum3.\*

41: em\_hum4.\*

42: gb\_pr4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	15.4	77.0	87210	7	AB011485 Arabidopsis

C	2	14.4	72.0	37304	1	MLCL373	AL035500 Mycobacte
c 3	14.4	72.0	4837	1	RSU81968	U81968 Rhodovulum	
c 4	14.4	72.0	358	14	G01494	G01494 fruit fly S	
c 5	14.4	72.0	3214	17	HBV131133	AJ131133 Hepatitis	
c 6	14.4	72.0	110000	34	CEY38E10_0	AL021149 Caenorhab	
c 7	14.4	72.0	110000	34	CEY57A10_3	Continuation (4 of	
c 8	14.4	72.0	125188	34	HSJ447F3	AL050348 Homo sapi	
c 9	14.4	72.0	163492	35	AC007241	AC007241 Homo sapi	
c 10	14.4	72.0	193347	35	AC007742	AC007742 Homo sapi	
c 11	14.4	72.0	14993	36	CEY57G9	Z83231 Caenorhabd	
c 12	14.4	72.0	3375	37	AF028726	AF028726 Trypanoso	
c 13	13.8	69.0	801	1	D84200	D84200 Escherichia	
c 14	13.8	69.0	17913	1	D90848	D90848 E.coli geno	
c 15	13.8	69.0	2664	1	HECUREASE	L25079 Helicobacter	
c 16	13.8	69.0	10036	2	AF000299	AF000299 Escherich	
c 17	13.8	69.0	43075	6	U88974	U88974 Streptococc	
c 18	13.8	69.0	4713	7	SCRKL197C	Z28197 S.cerevisia	
c 19	13.8	69.0	3479	7	YSCPAS1P	M58676 S.cerevisia	
c 20	13.8	69.0	99689	8	AC004557	AC004557 Genomic s	
c 21	13.8	69.0	76133	10	HSU71148	U71148 Human Xq28	
c 22	13.8	69.0	151316	11	AC005039	AC005039 Homo sapi	
c 23	13.8	69.0	112911	11	AC005865	AC005865 Homo sapi	
c 24	13.8	69.0	165516	11	AC005912	AC005912 Homo sapi	
c 25	13.8	69.0	35944	11	HSTCRB4H	AF009684 Homo sapi	
c 26	13.8	69.0	921	11	HSU70137	U70137 Human catlo	
c 27	13.8	69.0	216293	11	HTCRBCHR9	U66061 Human germ	
c 28	13.8	69.0	232650	11	U66061	AF072865 Rattus no	
c 29	13.8	69.0	1999	12	AF072865	AC006483 Homo sapi	
c 30	13.8	69.0	195764	34	CEY16B4	AL021148 Caenorhab	
c 31	13.8	69.0	187647	34	CEY16B4	Z98872 Caenorhabd	
c 32	13.8	69.0	123435	34	CEY61B8	AL033520 Homo sapi	
c 33	13.8	69.0	142347	34	HS349A12	AC007499 Homo sapi	
c 34	13.8	69.0	90106	35	AC007499	AC007779 Homo sapi	
c 35	13.8	69.0	178624	35	AC007779	AC007822 Drosophil	
c 36	13.8	69.0	50924	35	AC007822	AC005148 Drosophil	
c 37	13.8	69.0	95262	36	CEY42D1	Z81081 Caenorhabd	
c 38	13.8	69.0	32258	36	CEY42D1	Z92806 Caenorhabd	
c 39	13.8	69.0	42829	36	CEK10G4	U91634 Drosophila	
c 40	13.8	69.0	1689	36	DMU91634	U47881 Drosophila	
c 41	13.8	69.0	994	36	DMU47881	U49856 Drosophila	
c 42	13.8	69.0	1623	36	DMU49856	AF007813 Drosophil	
c 43	13.6	68.0	66194	35	AC007813	AF000869 Methanoba	
c 44	13.4	67.0	13699	1	AF000869	AF023464 Bacillus	
c 45	13.4	67.0	10513	1	AF023464		

#### ALIGNMENTS

RESULT	1	AB011485/c	87210 bp	DNA	PLN	05-FEB-1999
LOCUS	AB011485	Arabidopsis thaliana genomic DNA, chromosome 5, p1 clone: MXH1,				
DEFINITION	AB011485	complete sequence.				
ACCESSION	AB011485					
NID	G2924728					
VERSION	AB011485.1	GI:2924728				
KEYWORDS	HTG.					
SOURCE	Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1					
ORGANISM	Arabidopsis thaliana					
REFERENCE	1 (sites)					
AUTHORS	Nakamura, Y.					
TITLE	Structural Analysis of Arabidopsis thaliana Chromosome 5. V					
JOURNAL	Unpublished (1998)					
REFERENCE	2 (bases 1 to 87210)					
AUTHORS	Nakamura, Y.					
TITLE	Direct Submission					
JOURNAL	Submitted (02-MAR-1998) to the DDBJ/EMBL/GenBank databases.					

Fri Oct 1 16:47:14 1999

Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan (E-mail: ynakamu@kazusa.or.jp, Tel: +81-438-52-3935, Fax: +81-438-52-3934)

Location/Qualifiers

1..87210  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="5"  
/clone="MXH1"  
/clone\_lib="Mitsui PI"  
28245 a 15612 c 14898 g 28455 t

BASE COUNT  
ORIGIN

Query Match 77.0%; Score 15.4; DB 7; Length 87210;

Best Local Similarity 70.0%; Pred No. 70;  
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcdatrtgncraaacncc 20

||:||:|| ||:|| ||:||

Db 23648 GCGATGTGACCAACAGCC 23629

RESULT 2

MLCL373/c  
LOCUS MLCL373 37304 bp DNA BCT 26-FEB-1999  
DEFINITION Mycobacterium leprae cosmid L373.

ACCESSION AL035500

NID 94435688

VERSION AL035500.1 GI:4455688

KEYWORDS acyl-CoA carboxylase; acyl-CoA dehydrogenase; acyltransferase; ATP-dependent DNA helicase; biotin carboxylase; enoyl-CoA hydratase; galU; gated mechanosensitive ion channel; membrane protein; moeA; molybdenum cofactor; phosphoribosylaminoimidazolecarboxamide formyltransferase; phosphoribosylglycinamide formyltransferase; protein kinase; pseudogene; purH; RLEP; serine protease; succ; succinyl-CoA synthetase alpha subunit; succinyl-CoA synthetase beta subunit; synth; two-component response regulator; two-component system sensor; UTP-glucose-1-phosphate uridylyltransferase.

SOURCE

Mycobacterium leprae  
Mycobacterium leprae  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Mycobacteriaceae;  
Mycobacterium.

REFERENCE 1 (bases 1 to 37304)

Harris,D. and Taylor,K.

Unpublished

2 (bases 1 to 37304)

James,K.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.

Direct Submission

Submitted (17-FEB-1998) Mycobacterium leprae sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrrell@sanger.ac.uk Cosmids supplied by Dr. Stewart T. Cole, [3] Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France Requests for cosmids should be sent to Karin Eiglmeier (kei@pasteur.fr)

3 (bases 1 to 37304)

Eiglmeier,K., Honore,N., Woods,S.A., Caudron,B. and Cole,S.T.

Use of an ordered cosmid library to deduce the genomic organization of Mycobacterium leprae

Mol. Microbiol. 7 (2), 197-206 (1993)

93188700

Notes:

The Sanger Centre is funded to complete the sequence of M. leprae by the Heiser Program for Research in Leprosy and Tuberculosis of The New York Community Trust.  
Work in Paris is supported by the Heiser Trust, the Association Francaise Raoul Follereau and the Groupement de Recherches et des Etudes des Genomes (GIP-GREG).

Details of M. leprae sequencing at the Sanger Centre are available on the World Wide Web.  
(URL, <http://www.sanger.ac.uk/Projects/>)  
CDS are numbered using the following system eg MLCB33.01c. ML (M. leprae), CB33 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. All CDS over 100 codons have been analysed. Gene prediction is based on positional base preference in codons especially where there is an increase in the observed/expected third position G + C. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

FEATURES

source

Location/Qualifiers

1..37304

/organism="Mycobacterium leprae"

/db\_xref="taxon:1769"

/clone="cosmid L373"

1..755

/gene="MLCB373.01"

<1..755

/gene="MLCB373.01"

/note="MLCB373.01, probable ATP-dependent DNA helicase, partial CDS, len: >250 aa; similar to many e.g. PCRA\_BACST, Bacillus Stearotherophilus DNA helicase pcra (724 aa), fasta scores; opt: 402 z-score: 476.0 E(): 3.8e-19, 40.3% identity in 248 aa overlap and UVRD\_ECOLI (EMBL:M87049) E.coli DNA helicase II (720 aa) (39.8% identity in 196 aa overlap). Equivalent to Tr:P71561 (EMBL:Z79700) RV0949 (MTCY10D7.25c) M.tuberculosis probable DNA helicase (771 aa) (80.2% identity in 248 aa overlap)"

/codon\_start=3

/transl\_table=11

/product="probable ATP-dependent DNA helicase"

/protein\_id="CAB36663.1"

/db\_xref="PID:el388301"

/db\_xref="PID:94455689"

/db\_xref="GI:4455689"

/translation="PQELARDNLNELYSFAHEFSTEQANAAALAKSLHTPEDEVPD TGLAAFLKYSLSMDTQIPENNSGVVTLMTLHAAKGLFVFTVGTWEDGMLPHMR TLGDTTEISEERLAVGTRARQRLYSRAITRSSWGPIPLNPSRFUREIPPELID WRRSLTDSYSTPASGASRFGRVPSRRSGSKRALLVLPAGDRVTHDKYGLGRVEE VSGVGEAISLIDFGSGRIKLHMHAPVKL"

complement(1307..2338)

/gene="MLCB373.02c"

complement(1307..2338)

/gene="MLCB373.02c"

/note="MLCB373.02c, hypothetical protein, len: 343 aa; unknown function, similar to regions of many bacterial hypothetical proteins e.g. to the C-terminus of Tr:P74517 (EMBL:D90915) from Synecocystis sp. strain PCC6803 (715aa), fasta scores; opt: 446 z-score: 499.0 E(): 1.8e-20, 52.5% identity in 122 aa overlap. Equivalent to Tr:P71560 (EMBL:Z79700) RV0950c (MTCY10D7.24) M.tuberculosis hypothetical protein (332 aa) (71.0% identity in 331 aa overlap)"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein MLCB373.02c"

/protein\_id="CAB36664.1"



```

/note="LHII subunit beta"
/codon_start=1
/transl_table=11
/product="peripheral light harvesting complex subunit beta"
/protein_id="AAB59006.1"
/db_xref="PID:g1754677"
/db_xref="PID:g1754677"
/db_xref="GI:1754677"
/translation="MTDDMDKVPVTGLTAEAEVHKOLIDGTRVFGAIALFAHFLAA
IATPWLG"
2141..2332
/gene="pucA"
2141..2332
/gene="pucA"
/note="LHII subunit alpha precursor"
/codon_start=1
/transl_table=11
/product="peripheral light harvesting complex subunit
alpha precursor"
/protein_id="AAB59007.1"
/db_xref="PID:g1754678"
/db_xref="GI:1754678"
/translation="MNNAKMMLVVRKPTVGIPFLVACIAASFLVHMLVLVTGMWDY
YSGSFEAASLVSNATLLS"
2141..2293
/gene="pucA"
/note="11 amino acids are cleaved from the C-terminus of
LHII alpha subunit precursor to form mature protein"
/product="peripheral light harvesting complex subunit
alpha"
2490..3854
/gene="pucC"
2490..3854
/gene="pucC"
/function="essential for assembly of the peripheral light
harvesting complex"
/codon_start=1
/transl_table=11
/product="putative regulatory protein PucC"
/protein_id="AAB59008.1"
/db_xref="PID:g1754679"
/db_xref="GI:1754679"
/translation="MNRSLKMAVNRATVGRFLPFAEASEDLPLSRLLRLSMPQVS
VGMAMVLVGNVNRVNIIVLEVPASIVGINISLPLFAPFRALIGKSDTHKSA LGWR
RVPYIKGTLQWGFAPIMFALIVLSGOESAAGPEWIGILSAVSEFLVAGVHTV
QTVGLATDLPREDQPNVGLVMVLGMIVSALLFGMLEDFYHAKLIKVIQGA
ANATMVFNIAWKMEARDVRARQRLGDPSPREANGLETGPNARNRLWLWVLG
TLGFGSLVLEPFQGVQVLDMSVAATKLTAAVAGTLYGFAWASRVLSRGYDPHAMA
GWAAGVGLPFAFATITFSATIQSEPVFVTMMAGFGAGLFSHTLTATMRSAKPAQVG
LALGAWGAVQATSAGAGIAGGVFDRVIRRSATSSARRPYSLSTRSNWRCWSRPSS
SATG"
3899..3935
terminator 762 a 1541 c 1658 g 876 t
BASE COUNT
ORIGIN

Query Match 72.0%; Score 14.4; DB 1; Length 4837;
Best Local Similarity 68.4%; Pred. No. 2.4e+02;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcdatrtgncraaancnc 19
||:|||||:|||||
Db 1106 CGCATATGACCGAATACTC 1088

RESULT 4
G01494 358 bp DNA STS 27-FEB-1995
LOCUS fruit fly STS Dm1847 clone DS07826 T7.
DEFINITION G01494
ACCESSION G01494
NID 9684897
VERSION G01494.1 GI:684897
KEYWORDS STS sequence; primer; sequence tagged site.
SOURCE fruit fly vector-PAd10SacII The Pl library was made by D. Smoller

```

```

ORGANISM
REFERENCE 1 (bases 1 to 358)
AUTHORS Drosophila STS
TITLE Rubin, G.
JOURNAL Drosophila STS
COMMENT Unpublished (1994)

Contact: Berkeley Drosophila Genome Project
Berkeley Drosophila Genome Project
Primer A: GATCTGTTTCATGTTGTTCCGG
Primer B: TTGCCCATATGCGTATATACA
STS size: 191
PCR Profile:
Annealing: 58 degrees C PCR Cycles: 32
Protocol:
Template: Pl Library Pools
Primer: 1 uM each
dNTPs: 250 uM each
Taq Poly: 0.05 units/ul
Total Vol: 15 ul
Buffer:
MgCl2: 1.5mM
KCl: 50 mM
Tris-HCl: 50 mM
pH: 8.3
Gelatin: .001 %

The Pl library has been distributed to 16 regional sites. A list
of these sites is available from FlyBase, via anonymous ftp to
ftp.bio.indiana.edu in the file
flybase/allied-data/genome-projects/lbl/LBL.doc.

FEATURES
source 1..358
/organism="Drosophila melanogaster"
STS 4..4358
primer_bind 4..24
BASE COUNT 110 a 75 c 67 g 102 t 4 others
ORIGIN

Query Match 72.0%; Score 14.4; DB 14; Length 358;
Best Local Similarity 68.4%; Pred. No. 2.2e+02;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcdatrtgncraaancnc 19
||:|||||:|||||
Db 182 GCAATATGCGCAATACAC 200

RESULT 5
HBV131133/c 3214 bp DNA VRL 03-MAR-1999
LOCUS Hepatitis B virus, complete genome, strain J11.
DEFINITION AJ131133
ACCESSION 94468847
NID 94468847
VERSION AJ131133.1 GI:4468847
KEYWORDS C gene; core protein; P gene; polymerase; S gene; surface protein;
X gene; X protein.
SOURCE Hepatitis B virus.
ORGANISM Hepatitis B virus.
REFERENCE 1 (bases 1 to 3214)
AUTHORS Lu, L.C.
TITLE Direct Submission

```

in D. Hartl's lab (see Smoller et al., Chromosoma 100: 487). Chromosomal position was mapped by polytene chromosome in situ hybridization in the Hartl lab (see Hartl et al., PNAS 91: 6824). STS's were generated by sequencing the ends of the Drosophila insert in these Pl clones, by the W. Kimmerly, C. Martin, and M. Palazzolo lab at LBL.

Drosophila melanogaster  
Eucaryotae; Metazoa; Arthropoda; Tracheata; Insecta; Diptera; Brachycera; Cyclorhapha; Drosophilidae; Drosophila.  
Rubin, G.  
Drosophila STS  
Unpublished (1994)

Contact: Berkeley Drosophila Genome Project  
Berkeley Drosophila Genome Project  
Primer A: GATCTGTTTCATGTTGTTCCGG  
Primer B: TTGCCCATATGCGTATATACA  
STS size: 191  
PCR Profile:  
Annealing: 58 degrees C PCR Cycles: 32  
Protocol:  
Template: Pl Library Pools  
Primer: 1 uM each  
dNTPs: 250 uM each  
Taq Poly: 0.05 units/ul  
Total Vol: 15 ul  
Buffer:  
MgCl2: 1.5mM  
KCl: 50 mM  
Tris-HCl: 50 mM  
pH: 8.3  
Gelatin: .001 %

The Pl library has been distributed to 16 regional sites. A list of these sites is available from FlyBase, via anonymous ftp to ftp.bio.indiana.edu in the file flybase/allied-data/genome-projects/lbl/LBL.doc.

FEATURES  
source 1..358  
/organism="Drosophila melanogaster"  
STS 4..4358  
primer\_bind 4..24  
BASE COUNT 110 a 75 c 67 g 102 t 4 others  
ORIGIN

Query Match 72.0%; Score 14.4; DB 14; Length 358;  
Best Local Similarity 68.4%; Pred. No. 2.2e+02;  
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcdatrtgncraaancnc 19  
||:|||||:|||||  
Db 182 GCAATATGCGCAATACAC 200

RESULT 5  
HBV131133/c 3214 bp DNA VRL 03-MAR-1999  
LOCUS Hepatitis B virus, complete genome, strain J11.  
DEFINITION AJ131133  
ACCESSION 94468847  
NID 94468847  
VERSION AJ131133.1 GI:4468847  
KEYWORDS C gene; core protein; P gene; polymerase; S gene; surface protein;  
X gene; X protein.  
SOURCE Hepatitis B virus.  
ORGANISM Hepatitis B virus.  
REFERENCE 1 (bases 1 to 3214)  
AUTHORS Lu, L.C.  
TITLE Direct Submission

JOURNAL Submitted (27-NOV-1998) Lu L.C., Microbiology and Immunology,  
Leicester University, University Road, Leicester, LE1 9HN, UK

REFERENCE 2 (bases 1 to 3214)

AUTHORS Fujiyama, A., Miyanochara, A., Nozaki, C. and Yoneyama, T.

TITLE Point mutation in the S gene of hepatitis B virus for a d/y or w/r  
subtypic change in two blood donors carrying a surface antigen of  
compound subtype ady/r or adw/r

JOURNAL Unpublished

FEATURES

source Location/Qualifiers

1..3214

/organism="Hepatitis B virus"

/viation

/strain="JUL1"

/db\_xref="taxon:10407"

/lab\_host="human hepatocytes"

/note="Hepadnaviridae"

1..3214

/gene="p"

1..3214

/gene="S"

1373..1837

/gene="X"

1373..1837

/gene="X"

/codon\_start=1

/product="X protein"

/protein\_id="CAB38227.1"

/db\_xref="PID:el391948"

/db\_xref="PID:g4468848"

/db\_xref="GI:4468848"

/translation="MDARLCQDLPARDVLCRPVGAESGRPFPGALGPASPPD  
VPPDHGALSLRGLPVCAFSAGPCALRFTSARMETTVNAHNLKVLKHKTLGLSA  
MSTDLKLVKDFKCNFNEWELGEEVRLKVFVLGCRHKLVCSAPCNFTSA"

1813..2451

/gene="C"

1813..2451

/gene="C"

/codon\_start=1

/product="core protein"

/protein\_id="CAB38228.1"

/db\_xref="PID:el391949"

/db\_xref="PID:g4468849"

/db\_xref="GI:4468849"

/translation="MQLFHLCLISCSPTVQASKLGLWGMQMDIDPKYKFGASVEL  
LGFPSDFPSIRLDLTASLYREALESPHCSPHTALRQAILCWGELMNLATWVG  
SNLEDPASRELVSVMNMGKLRQLLWFLHISCLTFGRVTVLELVSVFGVIRTPPA  
VPPNAPILSLPETTVRRGRSPRRTPSPRRRSQSPRRRSQSRSSQC"

Join(2305..3214,1..1284)

/gene="S"

/note="read-through stop codon"

/codon\_start=1

/transl\_except="(pos:3083..3085,aa:OTHER)

/product="polymerase"

/protein\_id="CAB38229.1"

/db\_xref="PID:el391950"

/db\_xref="PID:g4468850"

/db\_xref="GI:4468850"

/translation="MPLSYQHFRKLKLLDEAGLEELPRLADEGLNRVRAEDNLNG  
NLNLSIPVTHKVGNTGLYSVFCFNPQWQTFPDHQLQEDIVDRCKQFVGLTVN  
ENRRLKMPARFYPNVTKLPDLGKIPYYPHVNVHYFTRHLYHLTWKAGLYKR  
ESTRSASFCGSPYSEODLQGLRVFOTSKRHGDKSPCPOSIGILPRSSVGCIOQSL  
RQSLGLOPAGOLAGROGSGSIRARVHPFMGTGVPEXSRHHCASFSSGF  
NHRQVKRAYSLISTSEGHSSGHAVELHFFPNSSRSQSGQFVLCWMLQFRNSPFC  
SFGSGNTVRWSPFAVNLQSLNLSLSDVSAAFYHLLPAAHPHLLVVG  
SGLSRLYSRLSSINNOHRTMNLNSCLSNLYMLLYKTYGRKHLHYSHPI  
ILGFRKIPMGVGLSPFLAQFTSAICSVRRAPPCFLAFSMDVGVLAGKSVKLSHSL  
YAAVTHFLSLGIHLNPKYTKRWGYSNEMGYVIGCLGTPODHIQVQIKMWKRLPV  
NRPIDMKVQRIVGLGLFAEPTCGYPALMPYACIOAKAFTSPTYKAFITKYL  
NLIPVARORGLQVFADATFTGWLAIQHGRMGKPLCLCRSLRNS"

Join(2712..3214,1..835)

/gene="S"

/codon\_start=1

/product="surface protein"

/protein\_id="CAB38230.1"

/db\_xref="PID:el391951"

/db\_xref="PID:g4468851"

/db\_xref="GI:4468851"

/translation="MWLIITSRDDIYTLFGRVSYIKESPHVAFPHAGHILGNKIY  
SMGWSKPRKMGNTLNVNPLGFFPDHQLDPAFKANDNDPDLNPKHKNWPDANK  
VGVGAFPGFIFSHGLGWSKAGQILTPVPAASLLASTIGKSRQFTPLSPPLRDT  
HPQAMQWNTTFTQTDQPRVRLYFPAGSSGTVSPAQNTVSAISSILSKTGDVPV  
NMENIASGLGLPVLVLAQAGFLITKILITPQSLDSWMTSLNFGTGPVLCGNSQOI  
SSHSPPCCPICPGYRMCILRRFIPLCLILLCLIFLLVLDYQGMPLVCPILPGSST  
TSTGPCKTCTTTPAGGTSMETSCCTKPTDNGCTCIPSSWAFAYLWEMASVRESWL  
SLLVFFQVFWGLSPVWLVSIVMMWYGFSLYNILSPFNPILLPIFLCLWYII"

BASE COUNT 734 a 842 c 728 g 910 t

ORIGIN

Query Match 72.0% Score 14.4; DB 17; Length 3214;  
Best Local Similarity 68.4% Pred. No. 2.4e+02;  
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 godatrtgncraanaanc 19  
||:||||:||||:||||:|

Db 910 GCAATGTGCCCAACACCC 892

RESULT 6  
CEY38E10\_0/c

WFCOMMENT

Sequence split into 4 fragments LOCUS CEY38E10 Accession AL021149

Fragment Name	Begin	End
CEY38E10_0	1	110000
CEY38E10_1	100001	210000
CEY38E10_2	200001	310000
CEY38E10_3	300001	393707

LOCUS CEY38E10 393707 bp DNA HTG 21-MAY-1999

DEFINITION Caenorhabditis elegans chromosome II clone Y38E10, WORKING DRAFT  
SEQUENCE, in unordered pieces.

ACCESSION AL021149

NID G4760249

VERSION AL021149.3 GI:4760249

KEYWORDS HTG; HTGS\_PHASE1.

SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans

REFERENCE 1 (bases 1 to 393707)  
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

AUTHORS Wallis, J.

TITLE Direct Submission

JOURNAL Submitted (21-MAY-1999) Nematode Sequencing Project, Sanger Centre,  
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,  
Washington University, St. Louis, MO 63110, USA. E-mail:  
jes@sanger.ac.uk or rwenematode.wustl.edu

COMMENT On May 7, 1999 this sequence version replaced gi:4455375.  
IMPORTANT: This sequence is unfinished and does not necessarily  
represent the correct sequence. Work on the sequence is in progress  
and the release of this data is based on the understanding that the  
sequence may change as work continues. The sequence may be  
contaminated with foreign sequence from E.coli, yeast, vector,  
phage etc. Order of segments is not known; 800 n's separate  
segments.  
\* NOTE: This is a 'working draft' sequence.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

FEATURES

source Location/Qualifiers

1..393707

/organism="Caenorhabditis elegans"

/db\_xref="taxon:6239"

/chromosome="II"

/clone="Y38E10"

BASE COUNT 117486 a 69904 c 68898 g 121367 t 16052 others

ORIGIN

Fri Oct 1 16:47:14 1999

```

Query Match      72.0%; Score 14.4; DB 34; Length 110000;
Best Local Similarity 68.4%; Pred. No. 2.8e+02;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcdatrtgncraaanacc 19
||:|||||:|||||
Db 24821 GCTATATGCCCCAAAAGCCC 24803

RESULT 7
CEY57A10_3
WPCOMMENT
Sequence split into 5 fragments LOCUS CEY57A10 Accession AL020986
Fragment Name      Begin      End
CEY57A10_0         1      110000
CEY57A10_1        100001    210000
CEY57A10_2        200001    310000
CEY57A10_3        300001    410000
CEY57A10_4        400001    469134
Continuation (4 of 5) of CEY57A10 from base 300001 (AL020986 Caenorhabditis elegans chr4)

Query Match      72.0%; Score 14.4; DB 34; Length 110000;
Best Local Similarity 68.4%; Pred. No. 2.8e+02;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcdatrtgncraaanacc 19
||:|||||:|||||
Db 34404 GCTATATGCCCCAAAARCCC 34422

RESULT 8
HSDJ447F3
LOCUS
DEFINITION Homo sapiens chromosome 20 clone DJ447F3, WORKING DRAFT SEQUENCE,
in unordered pieces.
ACCESSION AL050348
NID 95051848
VERSION AL050348.6 GI:5051848
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 125188)
AUTHORS Lloyd D.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jun 12, 1999 this sequence version replaced gi:5051342.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E. coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Unfinished: dj447F3 Contig_ID: 02039 acc=AL050348
Length: 23673 bp Unfinished: dj447F3 Contig_ID: 00738
acc=AL050348 Length: 8518 bp Unfinished: dj447F3 Contig_ID:
acc=AL050348 Length: 1653 bp Unfinished: dj447F3 Contig_ID:
01398 acc=AL050348 Length: 9114 bp Unfinished: dj447F3
Contig_ID: 01194 acc=AL050348 Length: 75081 bp Unfinished:
dj447F3 Contig_ID: 00234 acc=AL050348 Length: 1193 bp
Unfinished: dj447F3 Contig_ID: 02054 acc=AL050348 Length: 1156
bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
Location/Qualifiers
source

FEATURES --
Location/Qualifiers
source

```

```

source
1. 125188
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="DJ447F3"
BASE COUNT 32970 a 29506 c 26656 g 31253 t 4803 others
ORIGIN

Query Match.      72.0%; Score 14.4; DB 34; Length 125188;
Best Local Similarity 68.4%; Pred. No. 2.8e+02;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 cdatrtnccraaanacc 20
||:|||||:|||||
Db 122355 CTATGTGTCACAAATACCCC 122373

RESULT 9
AC007241
LOCUS
DEFINITION Homo sapiens clone NH0157F24, WORKING DRAFT SEQUENCE, 10 unordered
pieces.
ACCESSION AC007241
NID 94572725
VERSION AC007241.1 GI:4572725
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 163492)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 163492)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (07-APR-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 2104: contig of 2104 bp in length
2122: gap of unknown length
2123 4951: contig of 2829 bp in length
4952 4969: gap of unknown length
4970 9537: contig of 4568 bp in length
9538 9555: gap of unknown length
14435: contig of 4880 bp in length
14453: gap of unknown length
22123: contig of 7670 bp in length
22124 22141: gap of unknown length
22142 30343: contig of 8202 bp in length
30344 30361: gap of unknown length
30362 38164: contig of 7803 bp in length
38165 58106: contig of 19924 bp in length
58107 58124: gap of unknown length
58125 109742: contig of 51618 bp in length
109743 109760: gap of unknown length
109761 163492: contig of 53732 bp in length.
Location/Qualifiers
source
1. 163492
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NH0157F24"

```

```

BASE COUNT 46854 a 34129 c 35757 g 46590 t 162 others
ORIGIN

Query Match 72.0%; Score 14.4; DB 35; Length 163492;
Best Local Similarity 68.4%; Pred. No. 2.8e+02;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcdatrtgncraanacnc 19
||:||||| ||:|||||
Db 57501 GCGATATGTCACACAC 57519

RESULT 10
AC007742 193347 bp DNA HTG 05-JUN-1999
LOCUS Homo sapiens clone NH0357E16, WORKING DRAFT SEQUENCE, 15 unordered
DEFINITION pieces.
ACCESSION AC007742
NID 95001504
VERSION AC007742.1 GI:5001504
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 193347)
REFERENCE Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 193347)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (05-JUN-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2216: contig of 2216 bp in length
* 217 2234: gap of unknown length
* 2235 7707: contig of 5473 bp in length
* 7708 7725: gap of unknown length
* 7726 16249: contig of 8524 bp in length
* 16250 16267: gap of unknown length
* 16268 26418: contig of 10151 bp in length
* 26419 26436: gap of unknown length
* 26437 35911: contig of 9475 bp in length
* 35912 35929: gap of unknown length
* 35930 45721: contig of 9792 bp in length
* 45722 45739: gap of unknown length
* 45740 55585: contig of 9846 bp in length
* 55586 55603: gap of unknown length
* 55604 66636: contig of 11033 bp in length
* 66637 78150: gap of unknown length
* 78151 78208: gap of unknown length
* 78209 90637: contig of 12429 bp in length
* 90638 104304: contig of 13649 bp in length
* 104305 104322: gap of unknown length
* 104323 121509: contig of 17187 bp in length
* 121510 121527: gap of unknown length
* 121528 137177: contig of 15849 bp in length
* 137178 137193: gap of unknown length
* 137194 160149: contig of 22956 bp in length
* 160150 160166: gap of unknown length
* 160167 193347: contig of 33181 bp in length.

```

---

```

FEATURES
source
Location/Qualifiers
1..193347
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NH0357E16"
BASE COUNT 58295 a 37110 c 37135 g 60557 t 250 others
ORIGIN

Query Match 72.0%; Score 14.4; DB 35; Length 193347;
Best Local Similarity 68.4%; Pred. No. 2.8e+02;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcdatrtgncraanacnc 19
||:||||| ||:|||||
Db 121069 GCGATATGTCACACAC 121087

RESULT 11
CEF57G9 14993 bp DNA INV 23-NOV-1998
LOCUS Caenorhabditis elegans cosmid F57G9, complete sequence.
DEFINITION Z83231
ACCESSION Z83231
NID G1729610
VERSION Z83231.1 GI:1729610
KEYWORDS HTG
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 14993)
REFERENCE Steward, C.
AUTHORS Direct Submission
TITLE Submitted (09-DEC-1996) Louis, MO 63110, USA. E-mail:
jesssanger@ac.uk or rwnematode.wustl.edu
JOURNAL 2 (bases 1 to 14993)
REFERENCE Wilson, R., Alnsough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kersey, J., Kirsten, J., Laister, N.,
Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A.,
Saunders, D., Showkeen, R., Smaldon, N., Smith, A., Sonhammer, E.,
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Sproat, J. and Wohlman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
TITLE Nature 368 (6466), 32-38 (1994)
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
COMMENT Coding sequences below are predicted from computer analysis, using
predictions from Genefinder (P. Green, U. Washington), and other
available information.
For a graphical representation of this sequence and its analysis
see:-
http://webc.sanger.ac.uk/cgi-
bin/display?db=wormbase&class=Sequence&object=F57G9
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone F57G9.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone F57G9 is at 41466 in
sequence AL032647.

```

The start of this sequence (1..100) overlaps with the end of sequence AL032647.  
The end of this sequence (14893..14993) overlaps with the start of sequence AL023847.

FEATURES source

location: 1. .14993  
 /organism="Caenorhabditis elegans"  
 /db\_xref="taxon:6239"  
 /chromosome="II"  
 /clone="F57G9"  
 complement(2578. .5070)  
 /gene="F57G9.4"  
 complement(join(2578. .2785,3390. .3514,3672. .3794,  
 3836. .3992,4613. .5070))  
 /gene="F57G9.4"  
 /codon\_start=1  
 /protein\_id="CAB05752.1"  
 /db\_xref="PID:el347402"  
 /db\_xref="PID:g3877802"  
 /db\_xref="GI:3877802"  
 /db\_xref="SPIREMBL:O62269"  
 /translation="MLKINSNDIWLPIHFDYDTFNQLVLSVEFSLYLCAIYL  
 NIXYLKMFHNRNLIAPLPGIWFELIIGKLITAYRLKILNPGFELGVHIEWT  
 SDPRKLKVSNGLELILFGFQWIMYMETIIFGVLAIAVERVAVSLIENSTFE  
 LFIPLFTVTSQFSLSTSLALQYKQVPELAOLPMWICCPFSAMAYFFVKKNESGE  
 RETINRRRRHSVSQOQFVAKENLRALYLGTRLVFVLSGIALCGIGITAFYDLIPP  
 FCCHFVNEFLPHYLSCLTAIESVPQKNNEFREVSVLGRCLKIGRLKIESENAMEIQ  
 DSTKMGRETIDLIYFQQLADSWI"  
 complement(8130. .9673)  
 /gene="F57G9.2"  
 complement(join(8130. .8334,8393. .8517,8566. .8688,  
 8745. .8901,9210. .9673))  
 /gene="F57G9.2"  
 /codon\_start=1  
 /protein\_id="CAB05750.1"  
 /db\_xref="PID:el347400"  
 /db\_xref="PID:g3877800"  
 /db\_xref="GI:3877800"  
 /db\_xref="SPIREMBL:O62267"  
 /translation="MTHINSSSVYIWSVVFYKPELSKLILSIFELSSCILCGYIL  
 NLSFVMLKIQLFHKNLMFLVPLFAFWHELIIKGFITAVRLKIVNPGELGHTVF  
 WINDPDKLEVAGSGGLELILFGFQWIMYMETIIFGVLAIAVERVAVSLIENSTFE  
 ERIVIPILITISQLLSISIALITISIGPFLARLFPVICAPLSVLVFLFIKHTNOS  
 LLKICQNPKRTRIFTSQOQVKNLRALRGLRGLTVVWVIFYISGFGIALTFGLLI  
 PAGGCHITENELFLHPYICLTAMFSIPQWRDPFKSLPQWRRLAKIEQWTVRIIE  
 VNQNSSSVTDIYFRQLTESWT"  
 complement(10107. .11996)  
 /gene="F57G9.1"  
 complement(join(10107. .10311,10730. .10854,10907. .11029,  
 11297. .11453,11533. .11996))  
 /gene="F57G9.1"  
 /codon\_start=1  
 /protein\_id="CAB05749.1"  
 /db\_xref="PID:el347399"  
 /db\_xref="PID:g3877799"  
 /db\_xref="GI:3877799"  
 /db\_xref="SPIREMBL:O62266"  
 /translation="MIKNTGTSTFWLPTVYFNEPLNKLIVSIFELLSVLICGIIH  
 NLSIVMKSQILQLFHKNLMFLVPLFAFWHELIIKGFITAYRLKIVNPGELGHTVF  
 WINDPDKLEVAGSGGLELILFGFQWIMYMETIIFGVLAIAVERVAVSLIENSTFE  
 KRIVPILIIICQVLAIFMTFIVINRVHPILARLFIPLCFISFVAVLWVKNKKNK  
 LQKIQNPKRTRIFLTSQOQVKNLRALRGLRGLTVVWVIFYISGFGIALTFGLLI  
 FVCGHFVNEFLPHYLSCLTAIESVPQKNNEFREVSVLGRCLKIGRLKIESENAMEIQ  
 EDSTKMGRETIDLIYFQQLADSWI"  
 complement(12571. .13368)  
 /gene="F57G9.3"  
 complement(join(12571. .12996,13051. .13368))  
 /gene="F57G9.3"  
 /note="cDNA EST EMBL:T01304 comes from this gene"  
 /codon\_start=1  
 /protein\_id="CAB05751.1"  
 /db\_xref="PID:el347401"  
 /db\_xref="PID:g3877801"  
 /db\_xref="SPIREMBL:O62268"

CDS	/db_xref="GI:3877801"															
	/db_xref="SPTREMBL:O62268"															
	/translation="MGVGRILSYIYTFVLCIIACSYAWNITVFGNDIAQVASCIV															
	VIYIFLIILCPLPCIRPEEQSLQIQTFINQGFIAVTSIANVLLPYVMLLGP															
	NVSSIIILCYLLVDQLGFAVEMIHFAFSIQYTFPKTKILSNPKQYTIILAFSVIM															
	EFYHFYMCWNISNLMPINILCTAINIMEIPALIVAFYAINSDRTHFSRVKPGVGL															
	ELIEIRKWDITKARIWKINESPDEHVLHV"															
	Join(complement(AL023847..1:428..733),															
	complement(AL023847..1:102..376),complement(14855..14993))															
	/gene="Y57A10C.1"															
	/note="cDNA EST yk247g9.5 comes from this gene; cDNA EST															
	yk247g9.3 comes from this gene"															
	/codon_start=1															
	/protein_id="CAB05753.1"															
	/db_xref="PID:e1347403"															
	/db_xref="PID:g3877803"															
	/db_xref="GI:3877803"															
	/translation="MGVYGRVLSYIYFVLGNFINISYAWNITNHFCTDIAQVVPYIV															
	AIIPLVAVPVVLFIRRGKALPIOMFYFNQVIIAISAVFLSICLPMIMTPPIRSLIT															
	TLAVILHCPVPVLLCGISIEVHLVSPIQYTFPKTRLSNLQKFKISAFFILETVY															
	YITGRDLNHHPTNILLITAINIMTIPALFVAFYACNADSIHSSRVQSGGLIVTRKW															
	IMKGMWNLHDHGPDEHEDICD"															
BASE COUNT	5352 a	2542 c	2536 g	4563 t												
ORIGIN																
Query Match	72.0%; Score 14.4; DB 36; Length 14993;															
	Best Local Similarity 68.4%; Fred.No. 2.5e+02;															
Matches	13; Conservative 3; Mismatches 2; Indels 0; Gaps 0;															
QY	1	g	c	a	t	a	t	a	t	a						
Db	GCTATATGCCAAACCC 3112															
	: : : : : : : : :															
RESULT 12																
AF028726/c																
LOCUS																
DEFINITION	AF028726.1 GI:3057129															

```

/db_xref="GI:3057130"
/translation="MATTYEEFAAKLDRDAEFAKKMEONKRRFFADKPDEATLSPM
KEHYEFKMIQIHTDKFNKMHSEHFKAKFAELLEQQKNAQFPFGK"
1298..1576
/gene="KMP-11/2"
1298..1576
/gene="KMP-11/2"
/notes="membrane associated protein"
/codon_start=1
/product="kinetoplastid membrane protein 2"
/protein_id="AAC38991.1"
/db_xref="PID:g3057131"
/db_xref="GI:3057131"
/translation="MATTYEEFAAKLDRDAEFAKKMEONKRRFFADKPDEATLSPM
KEHYEFKMIQIHTDKFNKMHSEHFKAKFAELLEQQKNAQFPFGK"
1871..2149
/gene="KMP-11/3"
1871..2149
/gene="KMP-11/3"
/notes="membrane associated protein"
/codon_start=1
/product="kinetoplastid membrane protein 3"
/protein_id="AAC38992.1"
/db_xref="PID:g3057132"
/db_xref="GI:3057132"
/translation="MATTYEEFAAKLDRDAEFAKKMEONKRRFFADKPDEATLSPM
KEHYEFKMIQIHTDKFNKMHSEHFKAKFAELLEQQKNAQFPFGK"
2464..2742
/gene="KMP-11/4"
2464..2742
/gene="KMP-11/4"
/notes="membrane associated protein"
/codon_start=1
/product="kinetoplastid membrane protein 4"
/protein_id="AAC38993.1"
/db_xref="PID:g3057133"
/db_xref="GI:3057133"
/translation="MATTYEEFAAKLDRDAEFAKKMEONKRRFFADKPDEATLSPM
KEHYEFKMIQIHTDKFNKMHSEHFKAKFAELLEQQKNAQFPFGK"
BASE COUNT 866 a 766 c 737 g 1006 t
ORIGIN

Query Match 72.0%; Score 14.4; DB 37; Length 3375;
Best Local Similarity 68.4%; Pred. No. 2.4e+02;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gcdatrtgncraanaacnc 19
||:|||||:|||||
Db 3212 GCTATGACCAACACAC 3194

RESULT 13
D84200 801 bp DNA BCT 06-APR-1999
LOCUS Escherichia coli DNA for phosphomethylpyrimidine kinase, complete
DEFINITION cds.
ACCESSION D84200
NID g4589381
VERSION D84200.1 GI:4589381
KEYWORDS thid; phosphomethylpyrimidine kinase.
SOURCE Escherichia coli (strain:W3310) DNA, clone_lib:kohara 8F4.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (sites)
AUTHORS Mizote, T., Tsuda, M., Smith, D.D., Nakayama, H. and Nakazawa, T.
TITLE Cloning and characterization of the thid/J gene of Escherichia coli
encoding a thiamin-synthesizing bifunctional enzyme,
hydroxymethylpyrimidine kinase/phosphomethylpyrimidine kinase
JOURNAL Microbiology 145 (Pt 2), 495-501 (1999)
MEDLINE 99173753
REFERENCE 2 (bases 1 to 801)

```

```

AUTHORS Mizote, T.
TITLE Direct Submission
JOURNAL Submitted (23-MAR-1996) to the DDBJ/EMBL/GenBank databases. Tomoko
Mizote, Yamaguchi Prefectural University, Department of Food &
Nutrition, 3-2-1, Sakurabatake, Yamaguchi, Yamaguchi 753, Japan
(E-mail:tmizote@po.cc.yamaguchi-u.ac.jp, tel:0839-28-2489,
Fax:0839-28-2251)
FEATURES
Location/Qualifiers
source 1..801
/organism="Escherichia coli"
/strain="W3310"
/db_xref="taxon:562"
/clone_lib="Kohara 8F4"
/map="46 min"
1..801
/gene="thid"
1..801
/gene="thid"
/codon_start=1
/translation="phosphomethylpyrimidine kinase"
/product="phosphomethylpyrimidine kinase"
/protein_id="BAA76742.1"
/db_xref="PID:d1040495"
/db_xref="PID:g4589382"
/db_xref="GI:4589382"
/translation="MKRINALTAGTDPSGAGIQADLKTFSALGAYGCSVITALVAQ
NTRGVSVYRIEDFVAQDLSVFSVDRIOTTKIGMLAETDIVEVAERLQYQIONV
VLDIVLAKSGDPLLSAVATULRSLLPQVSLTPNLPFAALLDPHARTQCEMLE
QGRSLAMGCGAVLMKGGHLDQSPDLFTREGEQRTAPRIMTKNTHGTGCTLSAA
LAALPRHTNWADTVQAKSWLSALQAQDLEVGHGIGVPHHFAWW"
BASE COUNT 178 a 215 c 229 g 179 t
ORIGIN

Query Match 69.0%; Score 13.8; DB 1; Length 801;
Best Local Similarity 65.0%; Pred. No. 5.2e+02;
Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gcdatrtgncraanaacnc 20
||:|||||:|||||
Db 605 GCATTATGACCAAAACACC 624

RESULT 14
D90848/c 17913 bp DNA BCT 21-MAR-1997
LOCUS E.coli genomic DNA, Kohara clone #359(46.8-47.2 min.).
DEFINITION D90848 AB001340
ACCESSION
NID g1736814
VERSION D90848.1 GI:1736814
KEYWORDS Complete and shotgun sequencing; ISS; gata; gatA; gatB; gatC; gatY; gatZ;
yehA; yehB.
SOURCE Escherichia coli (strain:K12) DNA, clone_lib:Kohara lambda miniset
library clone:Kohara clone #359.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 17913)
AUTHORS Mori, H.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-1996) to the DDBJ/EMBL/GenBank databases.
Hirotsada Mori, NARA Institute of Science and Technology, Res. &
Edu. Center for Genetic Info.: 8916-5 Takayama, Ikoma, Nara 630-01,
Japan (E-mail:h-mori@etc.aist-nara.ac.jp, Tel:81-7437-2-5660,
Fax:81-7437-2-5669)
2 (sites)
REFERENCE 2 (sites)
AUTHORS Ikeda, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horiuchi, T.,
Ikenoto, K., Inada, T., Isono, K., Itoh, T., Kanai, K.,
Kasai, H., Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M.,
Kitakawa, K., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H.,
Motomura, K., Nakamura, Y., Nashimoto, H., Nishio, Y., Oshima, T.,
Saito, N., Sampei, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,
Yamamoto, Y. and Yano, M.

```





us-09-026-400-6.rge

Fri Oct 1 16:47:14 1999

-----



Fri Oct 1 16:47:14 1999

PA (SUMO ) SUMITOMO CHEM CO LTD.  
PI Mori S, Nakanishi H, Takahashi M;  
DR WPI; 98-439341/38.  
PT New nicotianamine aminotransferase protein and DNA - useful for  
PT enhancing iron absorption of plant cells  
PS Example 2; Page 7; 17pp; English.  
CC The primers V48149 and V48150 were used in the cloning of nicotianamine  
CC aminotransferase can be used in a plasmid to transform plant cells to  
CC produce cells with enhanced iron absorption, and it is implied [though  
CC not stated] that plants with improved resistance to iron deficiency  
CC chlorosis in calcareous soils can be regenerated from the transformed  
CC cells. The gene fragment can be used to detect, amplify and/or isolate  
CC nicotianamine aminotransferase genes. 2 G; 2 T;  
SQ Sequence 20 BP; 4 A; 6 C; 2 G; 2 T;

Query Match 77.0%; Score 15.4; DB 1; Length 20;  
Best Local Similarity 100.0%; Pred. No. 3.8;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcdatrtgncraaancnc 20  
|||||  
DB 1 GCDATRTGNCRAAANCNC 20

RESULT 3  
V48148/c  
ID V48148 standard; cDNA to mRNA; 1910 BP.  
AC V48148;  
DT 27-OCT-1998 (first entry)  
DE Nicotianamine aminotransferase 58148.62 molecular weight protein, gene.  
KW ds; nicotianamine aminotransferase; plant; iron absorption;  
KW iron deficiency chlorosis.  
OS Gramineae sp.  
FH Key Location/Qualifiers  
FT 76..1731  
FT CDS /\*tag= a  
FT /product= "Nicotianamine transferase"  
PN EP-860499-A2.  
PD 26-AUG-1998. 102891.  
PF 19-FEB-1998; JP-037499.  
PR (SUMO ) SUMITOMO CHEM CO LTD.  
PI Mori S, Nakanishi H, Takahashi M;  
DR WPI; 98-439341/38.  
DR P-PSDB; W61643.  
PT New nicotianamine aminotransferase protein and DNA - useful for  
PT enhancing iron absorption of plant cells  
PS Claim 4; Page 14-15; 17pp; English.  
CC The nicotianamine aminotransferase can be used in a plasmid to transform  
CC plant cells to produce cells with enhanced iron absorption, and it is  
CC implied [though not stated] that plants with improved resistance to iron  
CC deficiency chlorosis in calcareous soils can be regenerated from the  
CC transformed cells. The gene fragment can be used to detect, amplify  
CC and/or isolate nicotianamine aminotransferase genes. 368 T;  
SQ Sequence 1910 BP; 462 A; 534 C; 546 G; 368 T;

Query Match 69.0%; Score 13.8; DB 1; Length 1910;  
Best Local Similarity 65.0%; Pred. No. 61;  
Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 gcdatrtgncraaancnc 20  
|||||  
DB 1181 GTGATGTGCCCAACACTCC 1162

RESULT 4  
V03325  
ID V03325 standard; cDNA; 1253 BP.  
AC V03325;  
DT 15-APR-1998 (first entry)  
DE Sequence of clone 15, which encodes a TRAF2 binding protein.

KW Human tumour necrosis factor receptor-associated factor 2; TRAF2;  
KW TRAF-2 binding protein; NF-kappaB activity; NF-kappaB induction;  
KW intracellular signalling activity; acute hepatitis;  
KW autoimmune-induced cell death; ss.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT 2..1252  
FT CDS /\*tag= a  
FT /note= "no ATG start or STOP codon given"  
PN W09737016-A1.  
PD 09-OCT-1997.  
PF 01-APR-1997; IL0117.  
PR 26-AUG-1996; IL-119133.  
PR 02-APR-1996; IL-117800.  
PA (YEDA ) YEDA RES & DEV CO LTD.  
PI Boldin M, Kovalenko A, Malinin N, Mett I, Wallach D;  
DR WPI; 97-503101/46.  
DR P-PSDB; W42401.  
PT DNA encoding tumour necrosis factor receptor-associated factor  
PT binding molecule - used for modulation or mediation in cells of the  
PT activity of NF-kB  
PS Claim 4; Fig 5; 127pp; English.  
CC The present sequence is that of clone 15, which encodes a TRAF2  
CC binding protein. Clone 15 is a partial clone, which lacks most of  
CC its 5' end of the coding DNA sequence. A cDNA library prepared from  
CC B-cells was screened for proteins that associate with TRAF2, and  
CC the present sequence isolated. The clone 15 protein is capable of  
CC binding to at least amino acids 222-501 of TRAF2. The TRAF-2 binding  
CC proteins can be used for modulation or mediation in cells of the  
CC activity of NF-kappaB or any other intracellular signalling activity  
CC modulated or mediated by TRAF2. TRAF-binding proteins are especially used  
CC for prevention or treatment of pathological conditions associated with  
CC NF-kB induction, e.g. acute hepatitis, autoimmune-induced cell death,  
CC e.g. death of the beta Langerhans cells or the pancreas that results in  
CC diabetes, the death of cells in graft rejection, the death of  
CC oligodendrocytes in the brain in multiple sclerosis, and AIDS-inhibited  
CC the AIDS disease. The proteins are also useful for screening of ligands  
CC capable of binding to a protein, which are useful for modulating  
CC cellular activity modulated/mediated by TRAF2.  
SQ Sequence 1253 BP; 341 A; 321 C; 334 G; 256 T;

Query Match 67.0%; Score 13.4; DB 1; Length 1253;  
Best Local Similarity 70.6%; Pred. No. 96;  
Matches 12; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 gcdatrtgncraaanc 17  
|||||  
DB 207 GCTATGTGCCCGAACAC 223

RESULT 5  
V21210/c  
ID V21210 standard; DNA; 58407 BP.  
AC V21210;  
DT 10-NOV-1998 (first entry)  
DE Methanococcus jannaschii large circular extrachromosomal element.  
KW Methanococcus jannaschii; methanogenic archaea; circular chromosome;  
KW genome; autotrophic; extrachromosomal element; identification; ds.  
OS Methanococcus jannaschii.  
PN W09807830-A2.  
PD 26-FEB-1998. U14900.  
PF 22-AUG-1997; US-024428.  
PR (GENO-) INST GENOMIC RES.  
PA (UNIV ) UNIV ILLINOIS FOUND.  
PA (UJJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
PI Buit CJ, Smith HO, Venter JC, White OR, Woese CR;  
DR WPI; 98-169145/15.  
PT Complete genome sequence of methano-genic archaeon, Methanococcus  
PT jannaschii - useful in identification of M. jannaschii genome  
PT fragment

PS Claim 13; Page 585-600; 614pp; English.  
 CC The present sequence represents the large circular extrachromosomal  
 CC element sequence of the *Methanococcus jannaschii* circular chromosome. The  
 CC present invention describes *M. jannaschii* open reading frames from the  
 CC genome sequence. The invention also describes a computer based system  
 CC for identifying fragments of the *M. jannaschii* genome that are  
 CC homologous to target nucleotide sequences, comprising: (a) data storage  
 CC means comprising the nucleotide sequence of the 1664976, 58407 or 16550  
 CC bp sequence (see V21209, V21210 and V21211), or a nucleotide sequence at  
 CC least 99.9% identical to it; (b) search means for comparing a target  
 CC sequence to the nucleotide sequence of the data storage means to  
 CC identify a homologous sequence, and (c) retrieval means for obtaining  
 CC the homologous sequence. The method, which is based on whole genome  
 CC random sequencing of an autotrophic archaeon *M. jannaschii*, the genome  
 CC of which consists of 3 physically distinct elements, a large circular  
 CC chromosome (the 1664976 bp sequence given in V21209), a large circular  
 CC extra-chromosomal element (the 58407 bp sequence given in V21210), and a  
 CC small circular extra-chromosomal element (the 16550 bp sequence given in  
 CC V21211), can be used in the identification of *M. jannaschii* genome  
 CC fragment.  
 SQ Sequence 58407 BP; 20163 A; 9034 C; 7447 G; 21761 T;

Query Match 67.0%; Score 13.4; DB 1; Length 58407;  
 Best Local Similarity 66.7%; Pred. No. 1.7e+02;  
 Matches 12; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 2 cdattgncrraanacc 19  
 I:|||||:|||||  
 Db 51951 CAATATGCCCAATATAC 51934

## RESULT 6

V74830  
 ID V74830 standard; DNA; 3122 BP.  
 AC V74830;  
 DT 16-MAR-1999 (first entry)  
 DE Staphylococcus aureus contig SEQ ID #519.  
 KW Computer readable medium; vaccine; S aureus infection; immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 KW toxic shock syndrome; ds.  
 OS Staphylococcus aureus.  
 FH key Location/Qualifiers  
 FT misc\_feature 1021..1080  
 FT /\*tag= a  
 FT /note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 FT 2821..2880  
 FT /\*tag= b  
 FT /note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 FT EP-786519-A2.  
 PD 30-JUL-1997.  
 PF 07-JAN-1997; 100117.  
 PR 05-JAN-1996; US-009861.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC; Choi GH, Dillion PJ, Fannon MR, Kunsch CA.  
 PI Rosen CA;  
 DR WPI; 97-374922/35.  
 PT Polynucleotide(s) and proteins derived from *Staphylococcus aureus* -  
 PT stored on computer readable medium and used in the production of  
 PT anti-S aureus vaccines  
 PS Claim 1; Page 1446-1448; 3271pp; English.  
 CC This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the S aureus DNA sequences allows putative functions to be assigned so

CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against *S aureus* infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC S aureus in a sample. S aureus is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the S aureus DNA sequences contained on the  
 CC computer readable medium.  
 SQ Sequence 3122 BP; 1117 A; 400 C; 585 G; 899 T;

Query Match 67.0%; Score 13.4; DB 1; Length 3122;  
 Best Local Similarity 66.7%; Pred. No. 1.1e+02;  
 Matches 12; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 3 datrtgncrraanacc 20  
 I:|||||:|||||  
 Db 31 AATGTGTCCCAATATACCC 48

## RESULT 7

Q67186/c  
 ID Q67186 standard; DNA; 41 BP.  
 AC Q67186;  
 DT 22-MAR-1995 (first entry)  
 DE Primer for amplifying rev responsive element of HIV-1 HXB2 strain.  
 KW Immunisation; vaccine; therapy; prophylaxis; defective gene;  
 KW non-functional gene; template; antisense; ribozyme; bupivacaine;  
 KW HIV; human immunodeficiency virus; ss.  
 OS Synthetic.  
 PN WO9416737-A.  
 PD 04-AUG-1994.  
 PF 26-JAN-1994; U00899.  
 PR 26-JAN-1993; US-008342.  
 PR 11-MAR-1993; US-029336.  
 PR 15-JUL-1993; US-093235.  
 PR 21-SEP-1993; US-124962.  
 PR 21-SEP-1993; US-125012.  
 PA (CONE/) CONEY L R.  
 PA (MERY/) MERYA M J.  
 PA (WANG/) WANG B.  
 PA (WEIN/) WEINER D B.  
 PA (WILL/) WILLIAMS W V.  
 PA (ZURA/) ZURAWSKI V R.  
 PI Coney LR, Merya MJ, Wang B, Weiner DB, Williams WV;  
 PI Zurawski VR;  
 DR WPI; 94-263787/32.  
 DT Method for introducing genetic material into cells - utilises  
 PT polynucleotide function enhancer and nucleic acid free of  
 PT retroviral particles, e.g. HIV immunisation  
 PT Example 49; Page 110; 136pp; English.  
 CC Genetic material may be introduced into the cells of an individual  
 CC by (a) contacting the individual's cells with a polynucleotide  
 CC function enhancer (bupivacaine) and (b) administering to the cells  
 CC the nucleic acid molecule free of retroviral particles. Nucleic  
 CC acid molecules which are delivered to cells may serve as genetic  
 CC templates for proteins that function as prophylactic and/or  
 CC therapeutic immunising agents; replacement copies of defective,  
 CC missing or non-functional genes; genetic templates for therapeutic  
 CC proteins; genetic templates for antisense molecules or as genetic  
 CC templates for ribozymes. Two primers (Q67186, Q67187) were used  
 CC to amplify the rev responsive element (RRE) from a plasmid construct  
 CC comprising the RRE sequence from HIV-1 strain HXB2. The amplified  
 CC sequence was used in the construct designated pGAGPOL.rev2, used to  
 CC express HIV gag and pol genes.  
 SQ Sequence 41 BP; 10 A; 2 C; 17 G; 12 T;



CC and the gag open reading frame (ORF), part of the pol ORF and the rev  
 CC response element from HIV strain HXB2.  
 SQ Sequence 70 BP; 15 A; 8 C; 22 G; 25 T;

Query Match 66.0%; Score 13.2; DB 1; Length 70;  
 Best Local Similarity 70.6%; Pred. No. 82;  
 Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 4 atrtgncraanacncc 20  
 ||:||||:||||:||||  
 Db 47 ATATGCCCAAAACCCC 31

## RESULT 11

ID T32926/c  
 AC T32926 standard; DNA; 41 BP.  
 DT 12-FEB-1997 (first entry)  
 DE HIV-1 HXB2 rev response element (RRE) gene, PCR primer.  
 KW Primer; polymerase chain reaction; PCR; amplification; HIV-1;  
 KW human immunodeficiency virus type 1; HXB2; gag; pol; rev;  
 KW expression plasmid; induction; mucosal immunity; RRE;  
 KW response element; ss.  
 OS Synthetic.  
 PN WO9618390-A1.  
 PD 20-JUN-1996.  
 PF 15-DEC-1995; U16206.  
 PR 16-DEC-1994; US-357398.  
 PA (UYPE-) UNIV PENNSYLVANIA.  
 PI Ugen KE, Wang B, Weiner DB;  
 DR WPI: 96-300372/30.  
 PT Introduction of genetic material into mucosal tissue cells - by  
 PT topical or lavage admin. to induce mucosal immunity or treat disease  
 PS Example 14; Page 79; 91pp; English.  
 CC The primer pair T32926/27 was used for the PCR amplification of the  
 CC human immunodeficiency virus type 1 (HIV-1) HXB2 rev response  
 CC element (RRE) gene, to produce an amplification prod. which can be  
 CC used in the prepn. of a plasmid for the expression of the HIV-1  
 CC HXB2 gag and pol genes. The plasmid can be introduced into an  
 CC individual's cells by topical or lavage admin. to the rectal,  
 CC vaginal, urethral, sublingual or buccal mucosal tissue, to induce  
 CC mucosal immunity against the expressed protein and therefore  
 CC HIV-1.  
 SQ Sequence 41 BP; 10 A; 2 C; 17 G; 12 T;

Query Match 66.0%; Score 13.2; DB 1; Length 41;  
 Best Local Similarity 70.6%; Pred. No. 76;  
 Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 4 atrtgncraanacncc 20  
 ||:||||:||||:||||  
 Db 17 ATATGCCCAAAACCCC 1

## RESULT 12

ID T32927/c  
 AC T32927 standard; DNA; 70 BP.  
 DT 12-FEB-1997 (first entry)  
 DE HIV-1 HXB2 rev response element (RRE) gene, PCR primer.  
 KW Primer; polymerase chain reaction; PCR; amplification; HIV-1;  
 KW human immunodeficiency virus type 1; HXB2; gag; pol; rev;  
 KW expression plasmid; induction; mucosal immunity; RRE;  
 KW response element; ss.  
 OS Synthetic.  
 PN WO9618390-A1.  
 PD 20-JUN-1996.  
 PF 15-DEC-1995; U16206.  
 PR 16-DEC-1994; US-357398.  
 PA (UYPE-) UNIV PENNSYLVANIA.  
 PI Ugen KE, Wang B, Weiner DB;

DR WPI: 96-300372/30.  
 PT Introduction of genetic material into mucosal tissue cells - by  
 PT topical or lavage admin. to induce mucosal immunity or treat disease  
 PS Example 14; Page 79; 91pp; English.  
 CC The primer pair T32926/27 was used for the PCR amplification of the  
 CC human immunodeficiency virus type 1 (HIV-1) HXB2 rev response  
 CC element (RRE) gene, to produce an amplification prod. which can be  
 CC used in the prepn. of a plasmid for the expression of the HIV-1  
 CC HXB2 gag and pol genes. The plasmid can be introduced into an  
 CC individual's cells by topical or lavage admin. to the rectal,  
 CC vaginal, urethral, sublingual or buccal mucosal tissue, to induce  
 CC mucosal immunity against the expressed protein and therefore  
 CC HIV-1.  
 SQ Sequence 70 BP; 15 A; 8 C; 22 G; 25 T;

Query Match 66.0%; Score 13.2; DB 1; Length 70;  
 Best Local Similarity 70.6%; Pred. No. 82;  
 Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 4 atrtgncraanacncc 20  
 ||:||||:||||:||||  
 Db 47 ATATGCCCAAAACCCC 31

## RESULT 13

Q60144/c  
 ID Q60144 standard; DNA; 360 BP.  
 AC Q60144;  
 DT 16-MAR-1994 (first entry)  
 DE Human brain Expressed Sequence Tag EST02132.  
 KW Gene transcription product; genetic markers; tagging; in vivo;  
 KW transcriptions; mapping; locations; chromosomes; chromosomal; ss.  
 OS Homo sapiens.  
 PN WO9316178-A.  
 PD 19-AUG-1993.  
 PR 12-FEB-1993; U01294.  
 PR 12-FEB-1993; US-837195.  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
 PI Adams MD, Moreno RF, Venter CJ;  
 DR WPI: 93-272882/34.  
 PT Enriched oligonucleotides and corresp. sequences - used as  
 PT markers for human genes transcribed in-vivo, facilitate tagging  
 PT of most human genes  
 PS Example 4; Page 288; 500pp; English.  
 CC The Expressed Sequence Tag was isolated from a human brain cDNA  
 CC library as part of a large set of ESTs which can be used as markers  
 CC tagging of most human genes, for mapping locations of expressed genes  
 CC on chromosomes, for individual or forensic identification, for mapping  
 CC locations of disease-associated genes, for identification of tissue  
 CC type, and for prepn. of antisense sequences, probes and constructs.  
 CC EST02132 has a "poor" coding probability as evaluated using the  
 CC coding-region prediction program CRM. See also Q59041-Q61440.  
 SQ Sequence 360 BP; 89 A; 77 C; 98 G; 93 T;

Query Match 64.0%; Score 12.8; DB 1; Length 360;  
 Best Local Similarity 63.2%; Pred. No. 1.8e+02;  
 Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 2 cdartgncraanacncc 20  
 |:|:|:|:|:|:|:|:|  
 Db 115 CTATGTGTCGAGAGACCCC 97

## RESULT 14

Q70727  
 ID Q70727 standard; cDNA; 2018 BP.  
 AC Q70727;  
 DT 23-MAR-1995 (first entry)  
 DE TATA-binding protein-associated factor dTAFII60 cDNA.  
 KW TATA-binding protein associated factor; dTAFII60; ss; screening;



Fri Oct 1 16:47:14 1999

us-09-026-400-6.rng

---

Page 7

